Analysing locations power in large-scale mobility data

Lucas Santos de Oliveira, Pedro O S Vaz-De-Melo, Aline Carneiro Viana

Analysing locations power in large-scale mobility data

LUCAS SANTOS DE OLIVEIRA, Federal University of Minas Gerais, Brazil
PEDRO O.S. VAZ-DE-MELO, Federal University of Minas Gerais, Brazil
ALINE CARNEIRO VIANA, Inria Saclay - Ile de France, France

The pervasiveness of smartphones has modeled our lives, the social norms, and structure that dictate human behavior now directly influence the way individuals interact with network services and demand resources or content. From this scenario, identifying key locations in cities is central in human mobility investigation as well as for societal problem comprehension. In this context, we propose the first graph-based methodology in the literature to quantify the power of point-of-interests (POIs) over its vicinity by means of user mobility trajectories. Different from literature, we consider the flow of people in our analysis, instead of the number of neighbor POIs or their structural locations in the city. Thus, we modeled POI's visits using the multiflow graph model where each POI is a node and the transitions of users among POIs are a weighted direct edge. Using this multiflow graph model, we compute the *attract, support, and independence powers*. The *attract power* and *support power* measure how many visits a POI gather from and disseminate over its neighborhood, respectively. Moreover, the *independence power* captures the capacity of POI to receive visitors independently from other POIs. We tested our methodology on well-known University Campus mobility datasets and validated on Location-Based Social Networks (LBSNs) datasets from various cities around the world. Our findings show that in University campus: (i) buildings have low *support power* and *attract power*; (ii) people tend to move over a few buildings and spend most of their time in the same building; and (iii) there is a slight dependence among buildings, even those with high *independence power* receive user visits from other buildings on campus. Globally, we reveal that: (i) our metrics capture places that impact the number of visits in their neighborhood; (ii) cities in the same continent have similar independence patterns; and (iii) places with a high number of visitation and city central areas are the regions with the highest degree of independence.

CCS Concepts: • **Information systems → Mobile information processing systems**; *Data analytics*; **Spatial-temporal systems**; *Social networks.*

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Lucas Santos de Oliveira, Pedro O.S. Vaz-de-Melo, and Aline Carneiro Viana. 2018. Analysing locations power in large-scale mobility data. 1, 1 (February 2018), 28 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Internet devices and connections are growing faster than both the population and Internet users, as foreseen by Cisco [12]. Forecasts mention global mobile data traffic that will grow nearly twice as fast as fixed IP traffic from 2017 to 2022: Smartphones account for most of this growth [12, 13]. As a consequence, humans are immersed in a highly connected and ubiquitous cyber-physical context.

Authors' addresses: Lucas Santos de Oliveira, lsoliveira@uesb.edu.br, Federal University of Minas Gerais, Av. Antônio Carlos, 6627 - Predio do ICEx, Belo Horizonte, Minas Gerails, Brazil, 31270-901; Pedro O.S. Vaz-de-Melo, olmo@dcc.ufmg.br, Federal University of Minas Gerais, Belo Horizonte, Brazil; Aline Carneiro Viana, aline.viana@inria.fr, Inria Saclay - Ile de France, 1 rue Honore d'Estienne d'Orves. Campus de l'Ecole Polytechnique, Palaiseau, France, 91120.

XXXX-XXXX/2018/2-ART $15.00
https://doi.org/10.1145/1122445.1122456

Such pervasiveness of smartphones has remodeled our lives and made our real-life and virtual activities seamlessly merged together. i.e., most of our activities have now gone digital in some ways. In this context, mobile applications create a digital footprint that directly reflects our routines and whereabouts. Large datasets are currently being collected by various stakeholders to leverage this digital footprint in order to better learn our tastes, habits, and social lives, which also created new opportunities for research [18, 40].

This of course gave rise to deep studies of human activities and habits from such large sets of various data types [2, 43]. Social norms and structure dictating human behavior (e.g., mobility, interests) are now directly influencing the way individuals interact with the network services and demand resources or content. Many works address these opportunities by means of user mobility prediction and Point-of-Interest (POI) recommendation [17, 23, 43]. However, another fundamental area of study in traditional urban literature is the investigation of the way in which city neighborhoods become popular and how movement of citizens impacts the number of visitations in POIs [10, 15, 17, 18].

In fact, identifying key locations, which are places where persons spend a considerable amount of time during the day or which they visit frequently [34], is central to understand human mobility and social patterns. Such understanding can, in turn, inform solutions to large-scale societal problems in fields as varied as telecommunications, ecology, epidemiology, and urban planning [24]. In the literature, many works address human mobility as a graph-based approach [10, 35] that creates an opportunity for identifying important locations by means of power relationships, where one has the capacity to influence others or events.

In his seminal work on power-dependence relations, Richard Emerson [19] claims that power is a property of the social relation: i.e., saying "*X has power*" is vacant, unless we specify "*over whom*". Based on this claim, [8, 9] define power in the context of exchange networks, where the relationship in the network and their weights involve the transfer of valued items (i.e., information, time, money, energy). Different from centrality measures such as PageRank and closeness, which state that a node is central if it is connected with central nodes, Bozzo and Franceschet [9] consider that a node in a network is powerful if it is connected with powerless nodes. This notion of power leads to two assertions concerning the power of a node: (i) it is directly correlated with the number of its neighbors, and (ii) it is inversely correlated with the power of its neighbors [9]. In fact, the more ties a node has, the more powerful the node is. However, the second property, which is not habitual, characterizes power well: powerful nodes can impose their will on powerless nodes since the first has many other options to negotiate and the latter do not [8, 9].

Given the above context, in this paper, *we propose the first graph-based methodology in the literature to quantify the power of POIs by means of user mobility trajectories.* Different from [9], who infer the power of nodes from the sum reciprocal node degrees, *we infer the power of a node from its visiting flows* and according to the following three distinct approaches:

- First, *the attract power*, which is the capacity of a POI to receive people from its vicinity.
- Second, *the support power*, which is the capacity of a POI to disseminate people over its vicinity. In other words, given a large set of visits and mobility trajectories made by people, we calculate the potential impact [10], or influence, a POI has in its neighborhood. Imagine, for instance, a university campus and its impact in its nearby restaurants and bars when it is closed for summer vacations.
- Third and conversely, *the independence power*, which is the potential resilience a POI has to other POIs moving out from its neighborhood. Using the same example, replacing (or shutting down) restaurants and bars in the vicinity of the campus will probably not affect its visitations much.

Unlike traditional centrality metrics, *our metrics tackle the modeling of flows in exchange networks and identify places of power*. Besides, they clearly *distinguish among the three types of power a POI can assume*, namely power: to disseminate, to gather people, and to be indifferent to the flow of people visiting the POIs in the vicinity. It is worth mentioning that some literature studies analyzed the impact that a new business has on the local market ecosystem [10, 17] or the resilience in terms of business survival characteristics [16, 45]. Different from these works, *our work quantifies the impact of the POI in its vicinity individually, without category dependence*. Besides, our work measures the POI resilience by means of visitation independence, i.e., *it is invariant to the instability of vicinity visits*. Section 2 describes in detail the state-of-the-art.

Although we applied the methodology in the context of urban mobility, it can be used in other contexts, such as the analysis of the influence on social networks through the dissemination of information. Our metrics would help to identify who are the most influential people (*support power*) and who are the most passive people (*attract power*). Additionally, it can also be used to model message delivery problems in networks. The effectiveness of message delivery could be tested since our metrics identify and quantify the most powerful places to spread and to attract people.

In short, the main contributions of this work are fivefold:

- We propose a graph-based methodology named *Multiflow Graph Model* (MGM) to identify key locations, where each POI is a node and the transitions of users among POIs are a weighted direct edge. The weight of the edge is the number of transitions made by people who moved from one POI to the other. Therefore, from mobility user trajectories, we modeled the problem by means of Power relations among POIs.
- From this MGM approach, we propose three different influence measures: the *attract power* and *support power*, which are, respectively, the capacity of a POI to gather and to disseminate people, and the *independence power*, which is the POI capacity of receiving visitors independently from others POIs. Different from other metrics of Power, the *independence power*, *attract power*, and *support power* ranked as powerful distinct POIs (Section 3).
- We tested our methodology on well-known University Campus mobility datasets and validated on Location-Based Social Networks (LBSNs) datasets from various cities around the world (Section 4.1). Our findings show that in University campus: (i) buildings have low *support power* and *attract power*; (ii) people tend to move over a few buildings and spend most of their time in the same building; and (iii) there is a slight dependence among buildings, even those with high *independence power* receive user visits from other buildings on campus.
- Globally, we reveal that: (i) our metrics capture places that impact the number of visits in their neighborhood; (ii) cities in the same continent have similar independence patterns; and (iii) places with high number of visitation and city central areas are the regions with the highest degree of independence (Section 5).

In Section 6, we propose a practical application of our metrics in a case study of epidemic dissemination. Finally, we conclude and comment on the perspectives of our work in Section 7. *To the best of our knowledge, this is the first work in the literature that uses power relations, in terms of impact and independence, to infer the importance of a POI in its vicinity.*

2 RELATED WORK

The literature is rich in solutions that aim to leverage human mobility. The rise of mobile technologies and collective sensing in the last decade has contributed to the generation of large datasets that describe activity dynamics in cities and has created new research opportunities [18, 40]. Researchers addressed this challenging task using different data sources, such as location-based social networks and mobility traces.

Such mobility data sources have been used in the study of POI recommendations [47, 48]. The general idea of such works is to exploit the social connections and the favorite POIs of users to recommend new places to be visited. Ye et al. [47], for instance, explored user preference, social influence, and geographical influence to provide a POI recommendation service. Similarly, Zhang et al. [48] proposed a new approach called LORE to exploit the sequential influence of locations on users' check-in behaviors for location recommendations.

There are also works that analyze mobility data to predict, as accurately as possible, the future location of individuals and their friends [37, 41, 46]. D'Silva et al. [18], for instance, treated venue categories as proxies for urban activities to forecast the weekly popularity dynamics of a new venue establishment. Alternatively, Feng et. al. [20] proposed an attentional recurrent neural network model for predicting human mobility from lengthy and sparse trajectories. Moreover, Silveira et. al.[41] proposed a family of data-driven models, called MobHet, to predict human mobility using heterogeneous data sources. Comparably, Sadilek et. al. [37] explored the interplay between people's location, interactions, and their social ties. Then, they proposed a system that predicts the location and social ties in online social networks.

Another relevant research problem is the identification of relevant (or important) locations in mobility traces, i.e., personal places of interest (PPOIs): home, work, or any place where persons spend a considerable amount of time during the day or which they visit frequently [34]. Based on this problem, Pavan et al.[34] proposed a mapping scheme of POIs onto a feature space to identify those important locations. Alternatively, Isaacman et al. [24] proposed techniques based on clustering and regression for analyzing anonymized cellular network data, and to discern semantically-meaningful locations. Similarly, Cambe et al. [10] proposed a framework to examine the role of new businesses in their respective local areas. Using urban activity, they measure the impact, either positive or negative, that retail facilities have on each other. Contrastively, since POIs are sometimes difficult to be identified, Belcastro et al. [5] propose a technique that exploits the indications contained in geotagged social media items to discover regions of interest (RoI).

Different from this work, the method proposed in [10] is dependent on the POI category and uses a different impact percentage scale that varies around 1. Additionally, Cambe et al. [10] only analyze the impact of a location in its vicinity and do not identify the most important places (see Section 3.2). Finally, the authors do not take into account the independence of one location about its neighborhood (see Section 3.3).

Another way to infer the importance of POIs is by analyzing user transitions among them as an exchange network [4, 9]. Some studies addressed this problem as a relation of power, where the relationship in the network involves the transfer of valued items (i.e., information, time, money, energy)[4, 8, 9]. Moreover, it is advantageous to be connected to those who have few options. This type of relational power is endogenous concerning the network structures, meaning that it is a function of the position of the node in the network [8].

Additionally, the study of power has a long history in economics (in its acceptation of bargaining power) [31, 36] and sociology (in its interpretation of social power)[8, 14, 21]. In this work, we use the definition of power found in [8, 9]. However, we tackle this problem differently, taking into account the flow of people to identify powerful locations, not only the number of neighbors or its structural location in the network.

Summarized remarks: The literature on power inference is broad, special in economy and sociology domains. Although some interesting observations have emerged, the existing studies have several limitations: important places are not globally identified, the independence of places is not investigated, the position of places may influence relational power.

In a preliminary version of this work [32], we propose the *Multiflow Graph Model* and the metrics derived from it (*attract power*, *support power*, and *independence power*) and evaluate them in the Dartmouth campus dataset [27]. We here build on this prior effort by presenting a much more comprehensive investigation and offering:

- An extension of the methodology to incorporate a measure of uncertainty about the POI location.
- We evaluate our metrics in two more well-known University Campus mobility datasets and in three Location-Based Social Networks (LBSNs) datasets from various cities around the world.
- We delve deeper into the investigation of the results, assessing whether POIs ranked as powerful affect the visitation in the neighborhood and whether these locations change at different times of the day. Moreover, we investigated the different POIs independence patterns from different cities, and we compared regions with the highest proportions of POIs independence.
- Finally, we propose a practical application of our metrics in a case study of epidemic dissemination.

3 METHODOLOGY

In this section, we describe the *Multiflow Graph Model* where each location is a node and transitions of users between locations define the weighted directed edges in the graph. From this graph, we compute the *attract power*, *support power*, and *independence power* metrics, as described in the following. Notations used throughout this section are provided in Table 1

3.1 Multiflow Graph Model

The input necessary to construct the *Multiflow Graph Model* is a set of visits $\mathcal{C}$ over a set of POIs $\mathcal{P}$ made by a set of users $\mathcal{U}$. For simplicity, we organize the set of visits $\mathcal{C}$ into disjoint ordered sets $C_1, \cdots, C_{|U|}$, where $C_i = \{c^1, c^2, \cdots\}$ corresponds to all the visits c^j made by user u_i, in chronological order. A visit c^j is a tuple $< t, p >$, where t is the time the visit started and p is the POI where it took place. We denote by $t(c^j)$ and $p(c^j)$ the timestamp and the POI of *visit* c^j, respectively.

Then, from each ordered set C_i, we construct the corresponding set of *trajectories* S_i of user u_i. S_i contains all trajectories user u_i made during her social days, i.e., the distinct places u_i visited in chronological order. More formally, we transform every ordered set $C_i = \{c^1, c^2, \cdots\}$ into a set $S_i = \{S_i^1, S_i^2, \cdots\}$, where $S_i^k \in \mathcal{S}$ is a trajectory composed of a sequence of chronologically ordered POIs visited by user u_i, i.e., $S_i^k = \{p^1, p^2, \cdots\}$. This transformation is done by adding, in chronological order, the POI $p(c^j)$ of each visit $c^j \in C_i$ to a sequence $S_i^k \in S_i$.

The methodology we follow to construct the set of trajectories S_i is hereafter described and guarantees that any constructed trajectory has the following properties: (i) all transitions between consecutive POIs belonging to the same trajectory S_i^k happen in less than 6 hours (see item 1 hereafter); (ii) all visits of the same trajectory occur on the same social day, which begins at 6:00 and ends at 5:59 am of the next day (see item 2 hereafter); (iii) all POIs of a trajectory are distinct (see item 3 hereafter), i.e., each user trajectory S_i^k is a simple directed path with no cycles (see item 4 hereafter).

The process starts with the first user trajectory S_i^1, by making $k = 1$ and by adding the POI $p(c^1)$ of the first visit c^1 to S_i^1. Then, for each of the following visits c^j, we do the following:

1. **If $t(c^j) - t(c^{j-1}) \geq$ 6 hours, make $k = k + 1$ and add $p(c^j)$ to S_i^k;**
   We restricted $t(c^j) - t(c^{j-1}) < 6$ hours because Kotz et al. [26] suggested that around 90% of

Table 1. Notations

Notation	Description
$\mathcal{C}$	Set of visits
$\mathcal{P}$	Set of POIs
$\mathcal{U}$	Set of Users
c	A visit made by user u
p	POI visited by user u
C_i	All visits c made by user u, in chronological order
$t(c^j)$	Timestamp from visit c^j
$p(c^j)$	POI from visit c^j
$\mathcal{S}$	Set of trajectories
S_i	All trajectories of user u_i
κ	Trajectory signature
S_i^k	A trajectory composed of a sequence of chronologically ordered POIs visited by user u_i
H	Home. An artificial special location at the beginning and at the end of each user *trajectory* S_i^k
$n(\kappa)$	Number of trajectories with signature κ
S_κ	Ordered sequence of POIs defined by signature κ
$\|S_\kappa\|$	Size of S_κ
G	*Multiflow Graph Model*
V	Vertices that represent POIs in the graph G
E	Multi-Edges that represent user transitions between POIs
W	Edge weights
$w(p^i, p^j, \kappa)$	Edge weight from p^i to p^j in the trajectory κ
$N_{out}(p^i)$	Set of outgoing neighbors of node p^i
$N_{in}(p^i)$	Set of incoming neighbors of node p^i
$\mathcal{M}$	All messages posted in a given city present in the dataset
ϕ	GPS-latitude
λ	GPS-longitude
τ	Message timestamp
γ	the *range-impact* function
N_p	Set of all messages posted within a POI
d	Radius
σ	Sum of the *range-impact* harmonic means

user session durations are less than 6 hours. So, for differences $t(c^j) - t(c^{j-1})$ greater than or equal to $6h$, we consider that the user has started a new *trajectory*.

2. **Else If $t(c^j) \geq$ 6:00am and $t(c^{j-1}) \leq$ 5:59am, make $k = k + 1$ and add $p(c^j)$ to S_i^k;** We consider the user has started a new trajectory if two consecutive visits happen in different social days.
3. **Else if $p(c^j) == p(c^{j-1})$, do nothing and process POI $p(c^{j+1})$;** This ensures a trajectory has only distinct POIs.
4. **Else If $p(c^j) \in S_i^k$, make $k = k + 1$ and add $p(c^j)$ to S_i^k;** This prevents cycles to be formed. In special, if in a sequence of visits, the user returns to a POI already present in this sequence, we break this sequence and start a new trajectory from this latter visit.
5. **Else, add $p(c^j)$ to S_i^k;**

Some trajectories are dense and clearly represent what the user did in that day. On the other hand, other trajectories are composed by a single visit. In order to give the sense of flow to every trajectory, we added an artificial special location **H** at the beginning and at the end of each user *trajectory* S_i^k, simply called *Home*. Thus, each *trajectory* $S_i^k = \{p^1, \cdots, p^n\}$ is transformed to

$S_i^k* = \{p^0, p^1, \cdots, p^n, p^{n+1}\}$, where $p^0 \equiv \mathbf{H}$ and $p^n \equiv \mathbf{H}$. For simplicity, we will keep denoting each trajectory by $S_i^k = \{p^1, \cdots, p^n\}$, but with $p^1 = \mathbf{H}$ and $p^n = \mathbf{H}$ for all trajectories.

As the last step before constructing the *MGM*, we assign a *signature* κ to each trajectory $S_i^k = \{p^1, p^2, \cdots, p^n\}$, which is simply defined by the ordered sequence of POIs present in S_i^k, i.e., $\kappa(S_i^k) = p^1p^2(\cdots)p^n$. Note that trajectories with the same signature κ may exist within a set S_i and across different sets of users' trajectories S_i and S_j. All trajectories that have the same signature κ denote sequences of visits, or trajectories, over the exact same POIs. Additionally, we denote by $n(\kappa)$ the number of trajectories with signature κ. We also denote by S_κ the ordered sequence of POIs $p^1, p^2, \cdots, p^n$ defined by signature κ, which from now on we call *a signature trajectory* S_κ. The size of S_κ is denoted by $|S_\kappa|$. Finally, we denote by $\mathcal{S}$ the set of all *signature trajectories* S_κ.

From the *trajectories* in $\mathcal{S}$, we assemble the *Multiflow Graph Model* using the process described in Algorithm 1. For each trajectory $S_\kappa \in \mathcal{S}$, we traverse its sequence of locations in such a way that, for each pair of locations p^i and p^j, we create, in *line 4*, the edge $e(p^i, p^j, \kappa)$ between the POIs p^i and p^j. Additionally, in *line 5*, we assign to the edge the weight $w(p^i, p^j, \kappa)$ equal to the value from $n(\kappa)$.

Algorithm 1: Multiflow Graph Model - (MGM)

1 **Data:** Trajectory set $\mathcal{S}$
2 **for** $S_\kappa \in \mathcal{S}$ **do**
3   **for** $i \leftarrow 1$ **to** $\|S_\kappa\| - 1$ **do**
4         $e(p^i, p^j, \kappa)$;
5         $w(p^i, p^j, \kappa) = n(\kappa)$;
6   **end**
7 **end**

Figure 1 shows a figurative example of the *Multiflow Graph Model* that we represent as a MultiDiGraph $G(V, E, W)$, where: (i) vertices $p \in V$ represent POIs; (ii) the multi-edge arrows $e \in E$ represent the user transitions between POIs; and (iii) the weights $w(p^i, p^j, \kappa) \in W$ are represented on the arrows from p^i to p^j.

In the Figure 1, the trajectory signature κ is represented in colors, e.g., the trajectory from Academic building **A** to Restaurant building **R**, passing by Library building **L** that is colored in red.

Note that 10 users' trajectories start in **A** (red arrow), visit **L** and finish in **R**. Thus, the weight $w(p^i, p^j, \kappa) \in W$ represents the total number of trajectories from users who visited p^j just after visiting p^i, using the trajectory κ. Nodes Academic **A**, Library **L**, and Restaurant **R** represent campus buildings and nodes **H** represent the single artificial node *Home*. Although we represent **H** as multiple nodes in the figures, this is done only to ease the visualization.

3.2 Attract and Support power computation

In this section, we show how to compute the *support power* and *attract power* of POIs using the *Multiflow Graph Model* previously introduced. In other words, how to infer respectively, the capacity of a POI to disseminate and to receive people over and from its vicinity.

3.2.1 Attract power of POIs. The attract power of POIs are dependent on the out-degree distribution of edge weights. Therefore, we first compute the out-degree proportion for each weight $w(p^i, p^j, \kappa) \in W$ associated to each edge $e \in E$ as shown in Algorithm 2, line 3. This is done, in Eq. 1, by dividing the edge weight $w(p^i, p^j, \kappa)$ by the total sum of the edges weights w from the

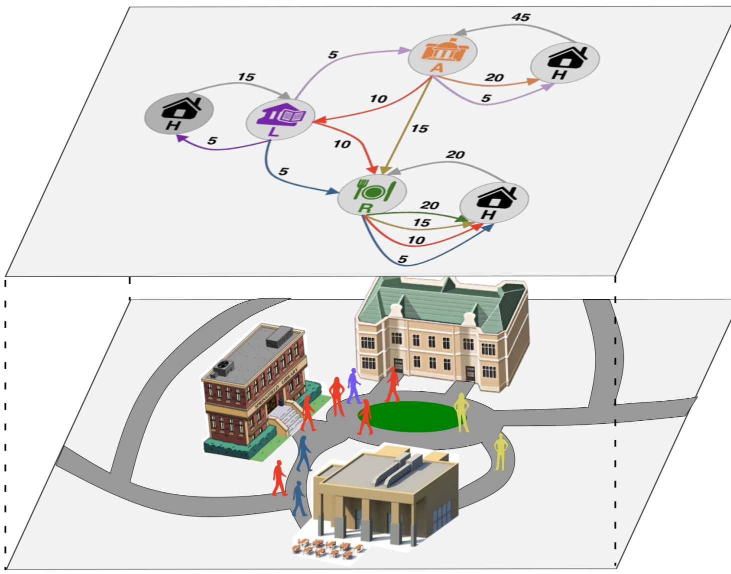


Fig. 1. Multiflow graph example (better seen in color).

outgoing neighbors of node p^i, given by the function N_{out}.

$$\frac{w(p^i, p^j, \kappa)}{\sum_{p^j \in N_{out}(p^i)} w(p^i, p^j, \_)} \tag{1}$$

We denote by $N_{out}(p^i)$ the set of outgoing neighbors of node p^i. Similarly, we denote by $N_{in}(p^i)$ the set of incoming neighbors of node p^i. Figure 2a shows the resulting out-degree proportion for each weight of the MultiDiGraph $G(V, E, W)$ example of Figure 1.

Algorithm 2: W_{out} edge weights out-degree proportion

Data: Edge weight set W, Outgoing neighbors function N_{out}
for $w(p^i, p^j, \kappa) \in W$ **do**
    $w_{out}(p^i, p^j, \kappa) \leftarrow \frac{w(p^i, p^j, \kappa)}{\sum_{p^j \in N_{out}(p^i)} w(p^i, p^j, \_)}$;
end

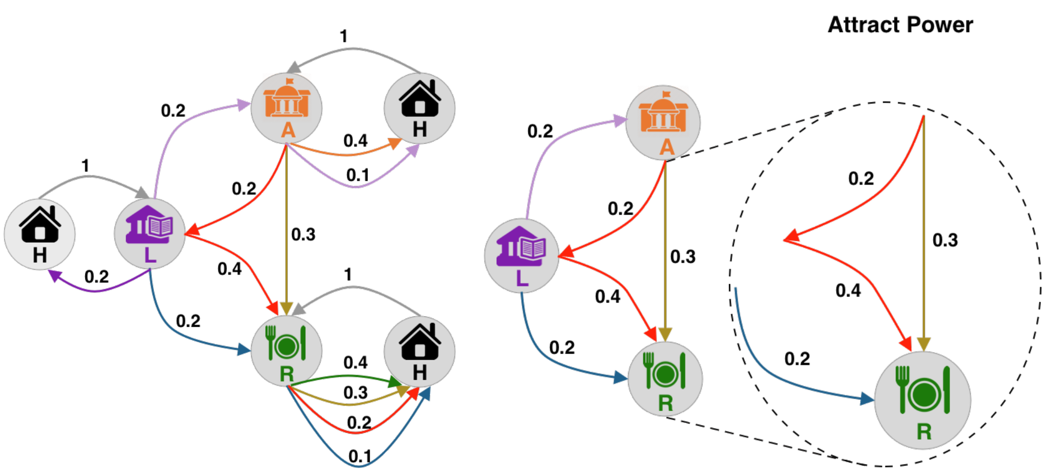
Then, after computing the out-degree proportions, we discard the Home **H** edge values as shown in Figure 2b. Finally, we compute the *attract power* of each node p^i according to the Algorithm 3, as follows. First, for each node $p^i \in V$, we initialize, in *line 3*, its *attract power* value with 0. Then, for each *trajectory* $S_\kappa \in \mathcal{S}$, we traverse all nodes p^i of this trajectory and sum cumulatively the proportions of edge weight $w_{out}(p^{i-1}, p_i, \kappa)$, given by Eq. 1, associated to the edge $e(p^{i-1}, p^i)$, *line 8*. At the same time, for each node p^i that we traverse, we add the cumulative total to its *attract power*, *line 9*.

The intuition of this algorithm is that POI p^i attracts a portion of the people who left the POI p^{i-1} and, throughout the trajectory, each POI p^i is indirectly responsible for this portion of visits in its predecessors: this is the reason why we use the cumulative sum.

Algorithm 3: Attract Power

Data: out-degree edge weight set W_{out}, Trajectory set $\mathcal{S}$, Vertices set V
for $p^i \in V$ **do**
  $attract[p^i] \leftarrow 0$;
end
for $S_\kappa \in \mathcal{S}$ **do**
  $tot \leftarrow 0$;
  for $i \leftarrow 2$ **to** $\|S_\kappa\| - 1$ **do**
    $tot+ = w_{out}(p^{i-1}, p^i, \kappa)$;
    $attract[p^i] + = tot$;
  end
end

Consider Figure 2b as an example. To compute the *attract power* of the restaurant **R**, we sum the weights from incoming edges from **A** and **L**: $0.3 + 0.4 + 0.2 + 0.2 = 1.1$. Note that we cumulatively sum the weights of the red edges from POI **A** to **R**. Similarly, to compute the *attract power* of the Academic building **A**, we use the unique income edge with a value of 0.2 (arrow purple). Finally, for library **L**, we use the unique value 0.2 (arrow red). In short, the *attract power* of POI **A** is the cumulative sum of importance from each POI in its incoming neighborhood.



(a) Out-degree distribution (b) Attract Power

Fig. 2. Attract Power of locations in *Multiflow Graph Model* (better seen in color).

It is important to mention that the *attract power* has two substantial meanings. First, it shows the capacity of the restaurant **R** to gather people and, consequently, how powerful POI **R** is. Second, if the location **R** closes its doors in this period, probably, 1.1 places will be impacted, because POI **R** will stop receiving visitors from these places.

3.2.2 Support power of POIs. Contrarily to *attract power*, the *support power* computation depends on the in-degree distribution of edge weights. Thus, we compute the in-degree proportion for each weight $w(p^i, p^j, \kappa) \in W$ associated to each edge $e(p^i, p^j, \kappa) \in E$ as shown in Algorithm 4, line 3. For each edge weight $w(p^i, p^j, \kappa) \in W$, we compute the in-degree proportion by dividing its weight by the total sum of the edges weights w from the incoming neighbors of node p^j, given by the function N_{in}, as shown in Eq. 2.

$$\frac{w(p^i, p^j, \kappa)}{\sum_{p^i \in N_{in}(p^j)} w(p^i, p^j, \_)} \tag{2}$$

Figure 3a shows the resulting in-degree distribution for each weight of the MultiDiGraph $G(V, E, W)$ of Figure 1.

Algorithm 4: W_{in} edge weights in-degree proportion

Data: Edge weight set W, Incoming neighbors function N_{in}
for $w(p^i, p^j, \kappa) \in W$ **do**
  $w_{in}(p^i, p^j, \kappa) \leftarrow \frac{w(p^i, p^j, \kappa)}{\sum_{p^i \in N_{in}(p^j)} w(p^i, p^j, \_)}$;
end

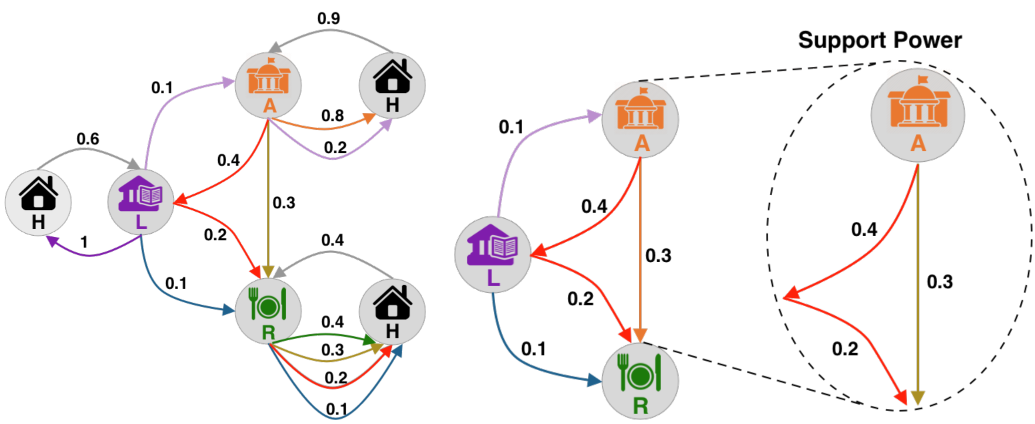
Then, after computing the in-degree proportions, we discard the Home **H** edge values as shown in Figure 3b. The *support power* of each node p^i is then computed according to Algorithm 5 described. More specifically, for each node $p^i \in V$, we first initialize its *support power* value with 0. Then, for each *trajectory* $S_\kappa \in \mathcal{S}$, we traverse all nodes p^i of this trajectory and sum cumulatively the proportions of edge weight $w_{in}(p^{i-1}, p^i, \kappa)$, given by Eq. 2, associated to the edge $e(p^{i-1}, p^i)$. At the same time, for each node p^{i-1} that we traverse, we add the cumulative total to its *support power*.

Note that different from *attract power*, to cumulatively calculate the *support power*, we traverse the path in reverse order, i.e. from the destination to the starting POI. The intuition of this algorithm is that POI p^{i-1} disseminates a portion of people who arrives in the POI p^i and, throughout the trajectory, each POI p^{i-1} is indirectly responsible for the portion of visits in its successors. This explains why we use the cumulative sum.

Algorithm 5: Support Power

Data: In-degree edge weight set W_{in}, Trajectory set $\mathcal{S}$, Vertices set V
for $p^i \in V$ **do**
  $support[p^i] \leftarrow 0$;
end
for $S_\kappa \in \mathcal{S}$ **do**
  $tot \leftarrow 0$;
  for $i \leftarrow \|S_\kappa\| - 1$ **to** 2 **do**
    $tot += w_{in}(p^{i-1}, p^i, \kappa)$;
    $support[p^{i-1}] += tot$;
  end
end

Consider the example shown in Figure 3b. To compute the *support power* of the academic building **A**, we sum the weight of all its outgoing edges: $0.2 + 0.4 + 0.3 = 0.9$. Note that we cumulatively sum the weights of the red edges from POI **A** to **R**. Similarly, to compute the *support power* of the library building **L**, we sum the outgoing edges' weights $0.1 + 0.2 + 0.1 = 0.4$. Finally, for restaurant **R**, the sum is 0, because there is no outgoing edge. In short, the *support power* of POI **A** is the cumulative sum of importance given by each POI in its outgoing neighborhood.



(a) In-degree distribution

(b) Support Power

Fig. 3. Support Power of locations in *Multiflow Graph Model* (better seen in color).

Similar to *attract power*, *support power* also has two substantial meanings. First, it shows the capacity of academic building **A** to disseminate people and, consequently, how powerful POI **A** is. Second, if the location **A** closes its doors in this period, probably, 0.9 places will be impacted, because **A** will stop sending visitors to these places.

3.3 Independence Power

The *Multiflow Graph Model* also enables us to compute another important metric that we call *independence power*, which shows the capacity of a POI to receive visitors independently from other POIs.

Using the special **H** location associated with each POI in the *MGM*, we can infer how many POIs have **H** as a starting point, in the *support power* case, and how many have **H** as an ending point, for the case of *attract power*.

More formally, Algorithm 6 shows how to compute the *independence power* of a POI. For each node $p^i \in V$, we sum to the variable *in*, *line 6*, the proportions of the incoming edges' weights $w(H, p^i, \kappa) \in W$, computed by dividing its weights by the total sum of the edges weights w from the incoming neighbors of node p^i, as shown in Eq. 3.

$$\frac{w(H, p^i, \kappa)}{\sum_{p^h \in N_{in}(p^i)} w(p^h, p^i, \_)} \tag{3}$$

Also, in *line 9*, we sum to the variable *out* the proportions of the outgoing edges' weights $w(p^i, H, \kappa) \in W$, computed by dividing its weights by the total sum of the edges' weights w from the outgoing neighbors of node p^i, as shown in Eq. 4.

$$\frac{w(p^i, H, \kappa)}{\sum_{p^j \in N_{out}(p^i)} w(p^i, p^j, \_)} \tag{4}$$

Finally, in *line 11*, the node p^i's *independence power* is given by the harmonic mean between the variables *in* and *out* values, as shown in Eq. 5.

$$2 * \frac{(in * out)}{(in + out)} \tag{5}$$

Similar to F1 Score, harmonic mean is appropriate for situations when the average of rates is desired.

Algorithm 6: Independence Power

Data: Edge weight set W, Vertices set V
for $p^i \in V$ **do**
    $independence[p^i] \longleftarrow 0$;
    in, out $\longleftarrow$ 0;
    for $w(H, p^i, \kappa) \in W$ **do**
        in += $\frac{w(H,p^i,\kappa)}{\sum_{p^h \in N_{in}(p^i)} w(p^h, p^i, \_)}$;
    end
    for $w(p^i, H, \kappa) \in W$ **do**
        out += $\frac{w(p^i,H,\kappa)}{\sum_{p^j \in N_{out}(p^i)} w(p^i, p^j, \_)}$;
    end
    $independence[p^i] = 2 * (in * out)/(in + out)$
end

Furthermore, the *independence power* value of a node ranges from 0 to 1, i.e., from totally dependent to totally independent from other POIs. Have a high *independence power* means that a POI receives visits regardless of the variation of visits in its neighborhood.

3.4 Assigning missing Points of Interest to geo-localized data

The selected datasets used in this study contain the exact information on which POI the data collection occurred, except the Twitter dataset [42]. For this latter, either it does not have the POI information or it is inaccurate [3, 5]. Therefore, we use the Bendler et al. [6] approach that uses a measure of uncertainty about the POI named ***range-impact***. Figure 4 illustrates a random sample of a map detail from San Francisco. The red dots represent positions of POIs, while the blue dots indicate Twitter messages. For each POI there is a neighborhood, indicated by the black circles, and the neighborhood distance/radius, indicated by the dashed lines. A Twitter message is assigned to a POI whenever its geo-tag is situated within the respective neighborhood. If the message was posted in an intersection area, we assign the message to the nearest POI.

More formally, hereafter, we define the available points of interest in a given city as the set of POIs $\mathcal{P}$, where each location $p \in P$ is represented by a 3-tuple $< c_p, \phi_p, \lambda_p >$, defined by Eqs. (6a) and (6b), where c_p is the category that the POI belongs, ϕ_p and λ_p define respectively, the GPS-latitude and GPS-longitude that marks the center of the POI's geographic coordinate. For this task, we use the OSMnx tool [7] that helps scholars to acquire, constructing, analyzing, and visualizing complex street networks.

$$P = \{p_1, p_2, ..., p_{\|P\|}\} \tag{6a}$$

$$p \in P \mapsto (c_p, \phi_p, \lambda_p) \tag{6b}$$

Furthermore, we define as the set $\mathcal{M}$ (Eq.(7a)), all messages posted in a given city present in the dataset. Each message $m \in \mathcal{M}$ defined in Eq. (7b) is represented by a 4-tuple $< u_m, \phi_m, \lambda_m, \tau_m >$ in

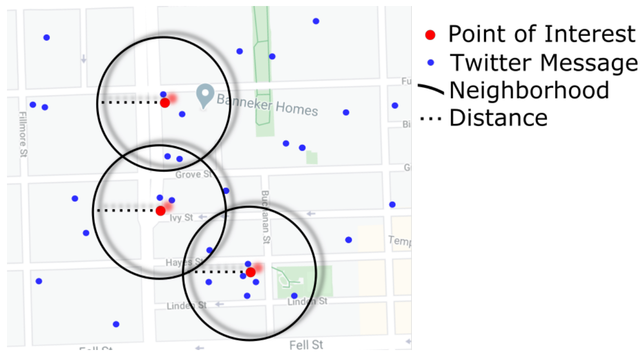


Fig. 4. Example of Point of Interest

which $u_m \in U$ is the user who posted the message, ϕ_m and λ_m refer respectively to the GPS-latitude and longitude where the message was posted. The attribute τ_m represents the timestamp in which the message was published.

$$\mathcal{M} = \{m_1, m_2, ..., m_{\|\mathcal{M}\|}\} \tag{7a}$$

$$m \in \mathcal{M} \mapsto (u_m, \phi_m, \lambda_m, \tau_m) \tag{7b}$$

Since we have set $\mathcal{P}$ and $\mathcal{M}$, we use the ***range-impact*** [6] to assign to a POI the messages that were posted within its neighborhood. As illustrated in Figure 4, we denote as N_p the set of all messages (blue dots) posted within the black circle around the reference POI (red dot), subject to a radius of d.

Then, for each message $m \in N_p$, we compute its ***range-impact*** on POI p using the *range-impact* function $\gamma(p, m)$ described in Eq. (8). The γ function takes as parameters the message m and POI p and returns the *range-impact* of message m on POI p. Internally, the γ function uses the coordinates of m and p to compute the haversine distance between the two locations. The haversine function is commonly used to calculate the spherical distance between two points on the Earth's surface in meters, given their latitude and longitude values.

$$\gamma(p, m) = \frac{1}{2} + \frac{1}{2}(\pi \frac{\text{haversine}(p, m)}{d}), \quad where \quad \gamma(p, m) \mapsto \mathbb{R}_0^+ \tag{8}$$

The *range-impact* function γ is modeled as a shifted and scaled cosine to fit the range of $[0, 1]$ on both x and y-axes, as illustrated in Figure 5. Messages that are quite close to the geographical origin of the POI p can still be seen as closely related to the locations and, thus, should be penalized on a negligible base. On the other hand, tweets that are far from the origin can be penalized with a substantially greater value but may still be related to the location itself [6]. This trade-off between a higher weight at short distances and a lower weight at far distances complies with the context stated by Tobler [44] in his first law of geography, i.e., *"[...] everything is related to everything else, but near things are more related than distant things"*.

Finally, we adapted the *MGM* algorithm 1 to take the *range-impact* into consideration. First, we transform each visit $c^j \in \mathcal{C}$ from tuple $< t, p >$ to 3-tuple $< t, p, \gamma >$. The γ is the *range-impact* value of visit c^j given by $\gamma(p, c^j)$, Eq. (8), where p is the nearest POI from visit c^j .

Next, at *line 5* of Algorithm 1, we initialize the edge weight $w(p^i, p^j, \kappa)$ with the σ value, as described in the Algorithm 7, instead of $n(\kappa)$. More specifically, Algorithm 7 takes as parameters

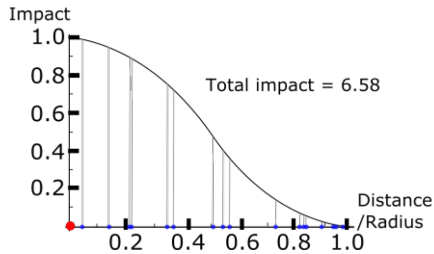


Fig. 5. Range Impact function.

the trajectory set $\mathcal{S}$, the trajectory signature S_κ, and the iteration index l. Then, it returns σ value, which is the sum of the *range-impact* harmonic means between the POIs p^i and p^j present in the user trajectories $S_u^k \in \mathcal{S}$, containing the signature S_κ. It is important to mention that the value of the edge weight still represents the flow of people who made the trajectory between the POIs p^i and p^j (as in the original *MGM*) but now weighted according to the haversine distance between messages and these POIs.

Algorithm 7: σ - MGM Edge Weight

```
Data: Trajectory set 𝒮, Trajectory signature S_κ, iteration index l
σ = 0 ;
for S_u^k ∈ 𝒮 do
    if S_u^k == S_κ then
        x = γ(p, c^l) ∈ S_u^k ;
        y = γ(p, c^m) ∈ S_u^k ;
        σ+ = 2 ∗ (xy)/(x + y) ;
    end
end
return σ ;
```

It should be pointed out that the Bendler et al. [6] approach presented in this subsection has some limitations. For example, depending on the size and shape of the POI, the neighborhood circle may not cover the total area of the POI which could lead to false negatives. Conversely, in the small POI area, the neighborhood circle is so big that aggregate messages that do not belong to the POI, leading to false positives. However, a large enough radius and the *range-impact* function above described minimizes these problems. Therefore, the results presented in Section 5 for Twitter dataset [42] using this methodology are parsimonious.

4 EVALUATION METHODOLOGY

4.1 Dataset

To evaluate and validate our proposed methodology, we use two groups of publicly available datasets, summarized in Table 2. The first group consists of campus mobility datasets from Dartmouth [27], KTH [33], and USC [25] universities. From now on, we call them *test datasets*, since they are well known and have been explored in several papers. The second group consists of Location Based Social Networks (LBSN) check-ins/posts from different cities around the world, which we call from now on, *validation datasets* since we use LBSN data in which we have less control over the location information of users and the area where the collection was performed. This group contains the following datasets: Gowalla [29], Weeplaces [29], and Twitter [42].

Table 2. Datasets Description.

	Trace source	Time/duration of trace	Start/End time	Granularity	Unique locations	City
Test datasets	USC	2006 summer semester	01/25/06 - 04/28/06	Building	137 buildings	Los Angeles
	Dartmouth	Fall 2003 and Winter 2004 terms	11/02/03 - 02/28/04	Building	61 buildings	Dartmouth
	KTH	16 months	01/01/14 - 04/30/15	Building	49 buildings	Stockholm
Validation datasets	Twitter	most of 2014-2017	03/14 - 04/17	Lat-Lon	> 80M tweets	11 Cities
	Gowalla	32 months	02/09 - 10/10	Lat-Lon	> 6M checkins	15 Cities
	Weeplaces	-	11/03 - 06/11	Lat-Lon	> 7M checkins	15 Cities

Test datasets:

According to Henderson et al. [22], Dartmouth College campus has over 190 buildings on 200 acres. In the Dartmouth [27] dataset, we focus on Syslog messages collected during the Fall 2003 and Winter 2004 terms, 17 weeks from 2 November 2003 to 28 February 2004, inclusive. In this work, we kept the building name and removed the AP number to eliminate cases where computers were associating and reassociating with several APs many times in succession.

The USC [25] data set was collected during 2003-2005 at the University of South California campus, where the number of WLAN users was over 4500. The USC trace has switch port location granularity which approximately corresponds to buildings on campus.

Finally, the KTH [33] dataset is the most recent available dataset (since 2019-07-01) and contains records of authenticated user associations to the wireless network of the KTH Royal Institute of Technology in Stockholm. The KTH wireless network provides coverage for buildings on one large and four small campuses located within the metropolitan Stockholm area. At the time of the trace collection (2014-2015) the university had around 18000 active students and employees, most of them accessing the wireless network via smartphones, laptops, and other portable devices.

Validation datasets:

Weeplaces [29] is a website that aims to visualize users' check-in activities in LBSNs. All the crawled data is originally generated in Foursquare. This dataset contains 7,658,368 check-ins generated by 15,799 users over 971,309 locations from various cities.

Gowalla [11] is a LBSN which had more than 600,000 users since November 2010 and was acquired by Facebook in December 2011. This dataset contains 6,442,890 check-ins made by 196,591 users from various cities. Finally, Twitter [42] that contains more than 80 Million of geotagged tweets from around 15 cities for most of 2014-2017.

4.2 Performance metrics

To evaluate the efficiency of the three here-above-presented power measures, we use the following metrics:

- Powerful building identification: we identify and rank the most powerful buildings for the different power metrics using the *test datasets* that provide a baseline for comparison.
- Relation among power metrics: we compare how our power metrics relate to each other, using the *test datasets* as a baseline.
- Neighborhood impact: we assess whether POIs ranked as powerful affect the visitation in the neighborhood and whether these locations change at different times of the day. For this task, we used some cities from the *validation datasets* which made it possible to find external information for validation.
- Independence power among cities: we investigated the different POIs independence patterns from different cities present in the *validation datasets*. Also, we compared regions with the highest proportions of POIs independence across different cities.

Also, when specified, we compare the performance of our power measures with the following literature ones:

- Betweenness centrality (betweenness): a graph measure based on shortest paths.
- Page rank (pagerank): computes a ranking of the nodes in a graph based on the structure of the incoming links. We use $alpha = 0.85$.
- In-degree centrality (in_dg_cen): computes a fraction of nodes connected to the incoming edges of a node.
- Out-degree centrality (out_dg_cen): computes a fraction of nodes connected to the outgoing edges of a node.
- Eigenvector centrality in (eigcen_in): computes the eigenvector for the largest eigenvalue of the adjacency matrix of a directed graph.
- Eigenvector centrality out (eigcen_out): computes the eigenvector for the largest eigenvalue of the adjacency matrix of a reversed directed graph.
- Bozzo and Franceschet power (bf_power): computes the sum reciprocal node degrees to identify the powerful nodes.

5 PERFORMANCE EVALUATION

In the following, we evaluate our *MGM* methodology to identify key locations as well as the three power measures. For this, the two sets of presented datasets, i.e., the test and the validation are used. Note that, as discussed in Subsection 4.1, the test set contains data which collection sampling is more regular, due to the way the data is collected (i.e., wireless network connectivity). On the other hand, validation datasets rely on the LBSN usability by users, and consequently, may have long temporal gaps. Because of its sampling properties and literature documentation, the set of test datasets allow a more precise verification of the results given by our measures.

5.1 Comparing the Campus Routine

Before showing the results for our proposed metrics, we perform a simple sanity check by examining the initial and final locations of the trajectories. For this task, we use the Dartmouth dataset which we consider representative for *test datasets*. We conjecture that if the trajectories represent daily routines on campus, then we expect a substantial amount to start and end at residential buildings. Table 3 shows the percentage of times each type of building started and ended a trajectory and its most ranked buildings. Observe that, as expected, most of the trajectories start and end at residential buildings. Also, it is expected that a large number of trajectories do not start and end at residential buildings, as some students do not live on campus or do not turn on their computers before leaving or after arriving home by the end of the day. Furthermore, in a similar analysis, Henderson et al. [22] and Kotz et al. [26] ranked most of these top buildings as the busiest on campus, given that these are communal areas visited by many, if not most students. These results lead us to conclude that our conjecture was correct and that our methodology captured the daily routine on campus.

Table 3. Buildings with highest starting and ending number of trajectories

Building type	**start**(%)	**end**(%)	**top start**	**top end**
Academic	**0.31**	**0.33**	Academic 2	Academic 2
Administrative	0.03	0.04	Administrative 1	Administrative 1
Athletic	0.02	0.02	Athletic 3	Athletic 3
Library	0.13	0.17	Library 2	Library 2
Residential	**0.43**	**0.36**	Residential 2	Residential 8
Social	0.08	0.08	Social 1	Social 1

5.2 Identifying the Powerful Buildings

Given that our approach captured the student transitions routine on campus, then in this section, we compute the *support power*, *attract power*, and *independence power* for all *test datasets*, as described in Section 3.2.

Furthermore, we compare our approach with the aforementioned metrics: Betweenness (betweenness), PageRank (pagerank), in-degree centrality (in_dg_cen), out-degree centrality (out_dg_cen), eigenvector centrality in (eigcen_in), eigenvector centrality out (eigcen_out) and Bozzo and Franceschet power [9] (bf_power).

Results are shown in Table 4, in which the meaningless names of buildings were changed for their functions. Note that the most powerful buildings are the social ones, such as libraries (Library 1, Library 2, Library 3, Library 4, Library 5), academics (Academic 2, Academic 1, Academic 4, Academic 10, Academic 14, Academic 16), and restaurants/social areas (Social 1, Social 3, Social 6). These buildings are hubs, i.e., they receive and disseminate students from and to all over campus, and are, in general, the busiest areas on campus [22].

Regarding the differences among the metrics, observe that in the Dartmouth dataset the *attract power* rank has two buildings that do not appear in the baselines' ranks: Academic 2 and Athletic 3. Intuitively, having a high *attract power* means that these buildings are responsible for receiving a large fraction of students from other campus buildings. Similarly, the *support power* rank has two buildings that do not appear in the baselines' ranks: Academic 2 and Athletic 3. Having high *support power* means that these buildings are responsible for a substantial fraction of people arriving at other campus buildings.

The *independence power* rank has all buildings that do not appear in the baselines' ranks. Having high *independence power* means that a substantial fraction of people has these buildings as unique destinations on campus. Additionally, these buildings with high *independence power* are, in the minority, administrative (Administrative 1), sports (Athletic 1, Athletic 2), and health (Health 1, Health 2, Health 3) buildings, and, in the majority, they are residential buildings (Residential 3, Residential 4, Residential 5, Residential 7), what corroborates with [26], since users spend more hours in residences than in other buildings.

It is important to mention that the first two positions of the rankings in the different datasets are very similar. As mentioned earlier, this is because these POIs, in general, are hubs. However, if we look at the entire rank for the different metrics, we can see that they are different. Moreover, the differences seen in our proposed power metrics are the consequence of an aspect that is not captured by the other baseline metrics: *the flow of people and the inter-dependence of flows among the buildings*. In this direction, note how the bf_power has a contrasting rank about our power metrics. Recall that the bf_power is computed using only the number of neighbors of the node, i.e., it ranks as powerful the buildings connected to the ones with the lowest degree. In the context of mobility transitions, it is not necessarily related to the interdependence among neighboring locations. Similarly, the metrics in_dg_cen and out_dg_cen only consider the degree of the vertices and rank as powerful the buildings with the highest degrees. However, the metrics PageRank, eigcen_in, and eigcen_out rank as important nodes the ones that are connected with other important nodes or, in the betweenness case, nodes that are most recurrent when calculating the shortest paths in the network. Nevertheless, they contrast with the definition of power in networks [9, 19], and do not capture the inter-dependence of flows among the buildings.

5.3 Relation among power metrics

In the previous Subsection 5.2, we ranked the buildings according to the three introduced power metrics in *test datasets*. From these ranks, we analyze how our metrics of power relate to each

Table 4. Top 5 College powerful buildings

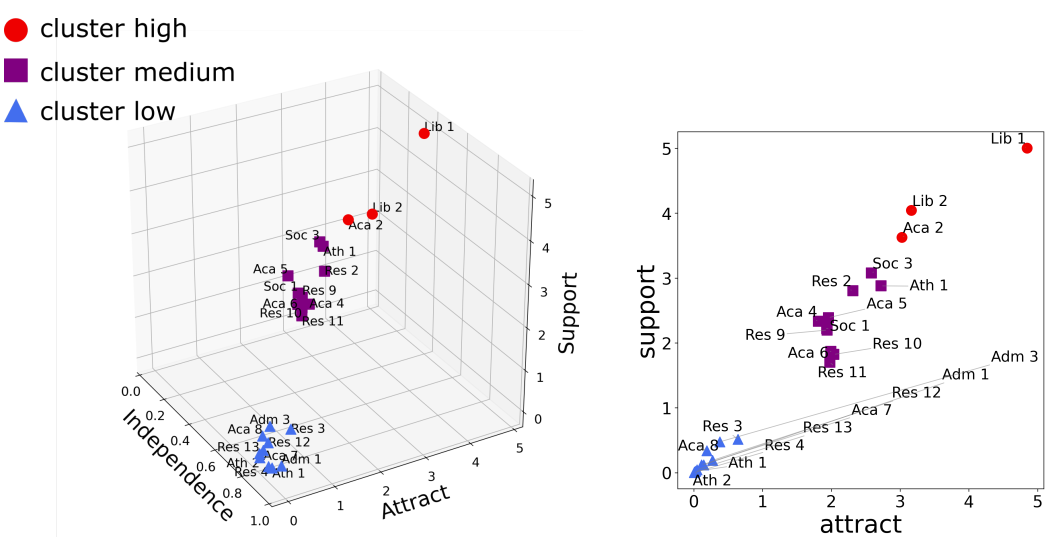
Dataset	Power	1st	2nd	3rd	4th	5th
Dartmouth	betweenness	Library 1	Library 2	Social 3	Social 1	Residential 2
	bf_power	Academic 1	Social 2	Residential 1	Library 1	Library 2
	in_dg_cen	Library 1	Library 2	Social 3	Social 1	Academic 3
	out_dg_cen	Library 1	Library 2	Social 3	Social 1	Academic 2
	eigcen_in	Library 1	Library 2	Social 3	Academic 3	Social 1
	eigcen_out	Library 1	Library 2	Social 3	Academic 3	Social 1
	pagerank	Academic 2	Library 1	Library 2	Athletic 3	Residential 2
	independence	Administrative 1	Athletic 1	Residential 4	Residential 3	Athletic 2
	support	Library 1	Library 2	Academic 2	Social 3	Athletic 3
	attract	Library 1	Library 2	Academic 2	Athletic 3	Social 3
USC	betweenness	Other 1	Library 3	Academic 4	Residential 5	Academic 5
	bf_power	Other 1	Library 3	Academic 4	Other 2	Administrative 2
	in_dg_cen	Other 1	Academic 4	Library 3	Residential 5	Academic 6
	out_dg_cen	Other 1	Academic 4	Residential 5	Library 3	Academic 6
	eigcen_in	Other 1	Academic 4	Residential 5	Library 3	Academic 7
	eigcen_out	Other 1	Academic 4	Library 3	Residential 5	Academic 7
	pagerank	Other 1	Library 3	Academic 4	Residential 5	Library 4
	independence	Health 1	Other 3	Health 2	Residential 6	Residential 7
	support	Other 1	Library 4	Library 3	Academic 6	Residential 5
	attract	Other 1	Library 3	Academic 6	Academic 4	Residential 5
KTH	betweenness	Academic 8	Academic 9	Academic 10	Academic 11	Administrative 2
	bf_power	Social 4	Academic 12	Academic 10	Academic 8	Academic 13
	in_dg_cen	Academic 10	Academic 14	Academic 15	Social 6	Academic 16
	out_dg_cen	Academic 10	Academic 14	Social 6	Academic 16	Academic 15
	eigcen_in	Academic 10	Social 6	Academic 15	Library 5	Academic 14
	eigcen_out	Academic 10	Social 6	Academic 15	Academic 14	Library 5
	pagerank	Academic 10	Social 6	Library 5	Academic 14	Academic 15
	independence	Other 3	Academic 17	Health 3	Other 4	Social 5
	support	Academic 10	Social 6	Academic 14	Academic 16	Academic 18
	attract	Academic 10	Social 6	Academic 14	Academic 15	Library 5

other for the most powerful buildings. For this task, we use the Dartmouth from *test datasets*, since we have the works of [22, 26] as a comparison baseline. Figure 6 shows the scatter plot of the top 10 ranked buildings from the Dartmouth dataset according to *attract power*, *support power*, and *independence power*. While the union set among the top 10 buildings for each metric totalizes 22 buildings, the intersection set results in 8 buildings. This shows that these metrics complement themselves, although, as expected, some natural overlap exists, especially between *attract power* and *support power*.

First, note in Figure 6a that there are three clusters of POIs: "cluster high" containing POIs with high *support power* and *attract power* ($\approx$ 3 to $\approx$ 5), and low *independence power* (< 0.4); "cluster medium" containing POIs with medium *attract power*, *support power* ($\approx$ 1 to $\approx$ 3) and low *independence power* (< 0.4); and "cluster low" containing POIs with low *attract power* and *support power* (≤ 1), and high *independence power* (≥ 0.4).

More specifically, observe in Figure 6b that "cluster high" contains *Lib 1*, *Lib 2* and *Aca 2* buildings that stand out as the three most powerful according with *attract power* and *support power*, with values ranging from $\approx$ 3 to $\approx$ 5. Intuitively, this means that these buildings have the potential to affect the visitation activity corresponding to the total amount of visits received by $\approx$ 3 to $\approx$ 5 other buildings. Furthermore, note that the highest-ranked buildings are, in the majority, social and library buildings, what is expected. According to [22], users spent less time in social and library than other buildings. Moreover, since these are communal areas, many users have those buildings in their routine trajectories, which explains their high power of disseminating and of gathering users.

Regarding the *independence power*, as we see in Figures 6a, the most powerful building is *Adm 1*, with *independence power* of 0.921. Intuitively, this means that 92.1% of its visits do not come from or go to other buildings, i.e., are completely independent. Also, note that the buildings in cluster high have moderate *independence power*, between 0.37 and 0.59. Conversely, the buildings in cluster low have small *attract power* and *support power* and large *independence power*, between 0.78 and 0.92. Additionally, these buildings with high *independence power* are, in the majority, residential buildings, what corroborates with [26], since users spend more hours in residences than in other buildings.



(a) top power metrics (b) *attract power* vs. *support power*

Fig. 6. Relation of power among metrics

In conclusion, the results shown in this section reveal that mobility in Dartmouth traces is very limited. People tend to move over a few buildings and spend most of their time in the same buildings. A similar conclusion was found in [22], which showed that 50% of users spend 74.0% of their time associated with a single access point. However, we identify a slight dependence among buildings as well as the tendency of people to be mostly stationary in a few buildings with short transit periods among them.

5.4 Neighborhood Visitation Impact

Previously, we identified in *test datasets* the most powerful buildings according to *attract power*, *support power*, and *independence power*. In this section, we use the *validation datasets* to investigate whether powerful buildings impact the visitation on POIs in their neighborhood. Table 5 shows the top 5 powerful buildings according to *support power* and *attract power* metrics for the cities of Pittsburgh - PA and Austin - TX. We chose these datasets due to the ease of finding events that occurred in the same periods and locations of the datasets collection which provided a baseline for the validation of the results.

Therefore, to assess the claim that a POI **A** impacts on POI **B**, we need to consider a counterfactual theory of causation of the form *"If **A** had not occurred, **B** would not have occurred"* [30]. However,

Table 5. Top 5 Cities powerful buildings

Dataset	Power	1st	2nd	3rd	4th	5th
Twitter - Pittsburgh	**support**	University	Parking	Church	Bank	Coffee shop
	attract	Conv. Center	University	Church	Bar	Ice cream shop
Weeplaces - Austin	**support**	Conv. Center	Hotel	Airport	Coffee shop	Supermarket
	attract	Conv. Center	Rock club	Hotel	Pub	Airport
Gowalla - Austin	**support**	Conv. Center	Hotel	Conv. Center	Conv. Center	Bar
	attract	Conv. Center	Hotel	Restaurant	Conv. Center	Hotel

Shrier and Platt [39] state that the counterfactual outcome may be dichotomous (e.g. the restaurant is open/closed) or continuous (e.g. if the number of people on campus varies, the number of visits to restaurants will vary). Indeed, the continuous counterfactual is a common method to investigate causality and it has been employed in most of the evidence-based sciences, where modifications in a variable change the outcome during a natural experiment. These evidence-based sciences include medicine [38], economics [49], and information systems [6, 28]. Hence, to estimate the impact of a POI on its vicinity, we need an exogenous variation in the availability of such POI. Such a variation is provided by events that occurred at the POI, which can be used as counterfactual to show that the number of visits in the neighborhood increases (decreases) if the POI is opened (closed).

To address this problem, we first search for POIs that were ranked as powerful using the *support power* and *attract power* metrics in *validation datasets*. Then, we identify which relevant events occurred at these POIs. Next, for each POI, we obtained the distributions of the number of visits in its vicinity during these events ($dist_{event}$) and compared it to the distribution of visits in a close period ($dist_{close}$). For this comparison, we use the two-sample Kolmogorov-Smirnov (KS) test, a statistic that quantifies a distance between the empirical distribution of two samples. The null hypothesis is that the two samples are drawn from the same distribution (p-value >= 0.05) or if they are drawn from different distributions (p-value < 0.05).

Our hypothesis:
H1: there are changes in the number of visits.
H0: there are no changes in the number of visits.

If the p-value of the KS test is greater than 0.05, then we have no evidence that changes in the number of visits on the POI affected their vicinity, i.e., the distributions $dist_{event}$ and $dist_{close}$ are drawn from the same distribution and we cannot reject hypothesis H0. Otherwise, we accept the hypothesis H1.

Applying the aforementioned methodology in *validation datasets*, we obtained for the city of Pittsburgh - PA, in the Twitter dataset, that Point Park University was ranked as powerful using our metrics. Moreover, as counterfactual, we look at the academic calendar from the interval comprising the dataset and identify 2 periods in which the university paralyzed its academic activities. The first period was thanks-given recess from Nov 16 to 30, 2014 and the second was the winter recess from Dez 13, 2014 to Jan 05, 2015. Then, for thanks-given recess, we compared the distribution of visits in the neighborhood with the academic activities period from Nov 1 to 16, 2014. For this event, we obtained the p-value from the KS test equals to 0.0005, which rejects the H0 null hypotheses, indicating changes in the number of visits. A similar result was found in winter recess which we compared with the return to academic activities period from Jan 06, 2015, to Feb 13, 2015. Also, we obtained the p-value equals to 2.0e-13, which rejects the H0 hypothesis. Both examples indicate thus, that changes in the activities of the university during thanks-given and winter recess affected the number of visits in its neighborhood.

A compared result was found for the city of Austin - TX, in the Weeplaces dataset. Our metrics ranked Austin Convention Center as powerful. Then, we searched for the main events that occurred at this location to identify a counterfactual. We found the SXSW conference, an annual conglomeration of parallel film, interactive media, music festivals, and conferences organized jointly that take place in mid-March. Then, for each year that comprises our dataset (2009, 2010, and 2011), we compared the distribution of visits in the neighborhood of Austin Convention Center during the SXSW week with the week before the event. Thus, for each year, we obtained from the KS the respective p-values (0.03, 1.7e-83 and 5.0e-39) which indicate changes in the number of visits in the Austin Convention Center neighborhood, rejecting the H0 hypothesis. This is an interesting result since the impact of the SXSW on the neighborhood occurred over the three different years. Moreover, we also identified the same location as powerful in the Gowalla dataset for the year of 2010. Similarly, we apply the same methodology and obtain the p-value for the KS test equals 2.0e-21, which also rejects the H0 hypothesis. Therefore, *we provided for three different datasets, two different cities, and three different years examples of counterfactual for the claim that a POI **A** impacts the visitation on POI **B** i.e. "If A varies, B also varies".*

5.5 Neighborhood Impact in Different Periods

We showed that POIs ranked as powerful by our metrics actually impact the number of visits in their neighborhood. However, one question that arises is whether this impact occurs throughout the day or at specific periods. To answer this question, we first aggregated the hours of the day over eight-hour time windows, corresponding to three periods of the day: morning (06:00-14:00), afternoon (14:00-22:00), and night (22:00-06:00). Then, we applied our methodology to the data from the cities of San Francisco and Pittsburgh, considering the three distinct periods of the day. These two cities were chosen because they have the largest amount of data. Finally, for each city, we sampled the most powerful POI by metric and by the period of the day.

Table 6. Powerful places by period of the day

	San Francisco - Weeplace		**Pittsburgh - Twitter**	
	attract	support	attract	support
06:00-06:00	Airport	Airport	Convention Center	University
06:00-14:00	Fast-Food	Airport	Convention Center	Residential Parking
14:00-22:00	Store	Airport	Cafe	Hospital
22:00-06:00	Airport	Airport	University	Commercial Parking

The results are shown in Table 6. For the city of San Francisco, the International Airport was ranked as powerful to disseminate and to gather people. However, during the morning, a fast-food restaurant stood as the one that most attracted people and, during the afternoon, a large electronics store. Differently, for the city of Pittsburgh, when we do not separate by the period of the day, the Convention Center was ranked as the highest *attract power* and a University was ranked as the highest *support power*. However, when we separate by the period of the day, each period has a different POI ranked as powerful. During the morning the one with the highest *attract power* was the Convention Center and, the one with the highest *support power* was a residential parking. Also, during the afternoon a coffee shop stood as the most powerful to attract people, and a hospital, the one that disseminated the most. Finally, during the night, a university had the highest *attract power* and a commercial parking had the highest *support power*. These results show that in different cities there are totally different behaviors that vary according to the period of the day. Some places are very important in the city and act as hubs, disseminating and attracting people, others are powerful depending on the time of day. These behaviors in different cities ask for a deeper investigation of human dynamics in urban areas, through the comparison of similar patterns across different cities [40].

5.6 Independence Power Among Cities

In section 5.2 we show that even powerful POIs have a certain degree of independence from their neighborhood in campus scenario's datasets. However, is this same behavior found in urban scenario's datasets? More, is there any difference in the pattern of independence of POIs for different cities?

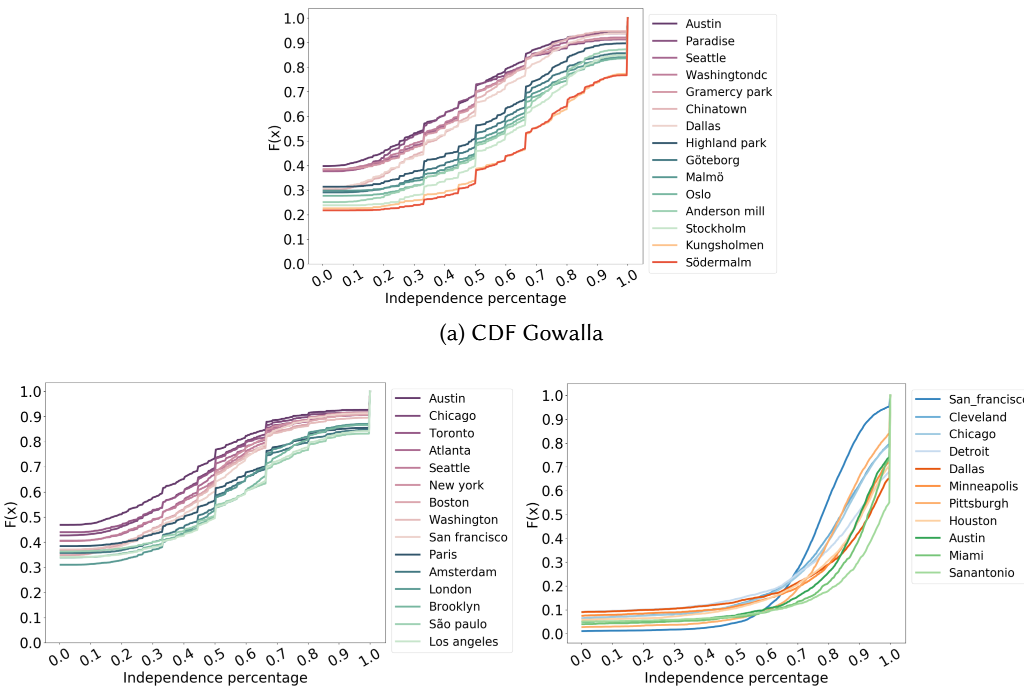
To answer these questions we show in Figure 7 the cumulative distribution functions (CDFs) of POI independence for each city from *validation datasets*. Figure 7a shows these CDFs for the Gowalla dataset, where we can easily visualize three distinct groups of cities. The first group is formed by U.S. cities that have a lower degree of independence compared to the other groups. About 50% of POIs have at least 30% independence. The second group is formed by European cities and has greater independence from POIs about U.S. cities. About 50% of POIs have at least 50% independence. The third group is formed by two islands around the city of Stockholm. The two islands have a similar pattern between them and are different from the city of Stockholm, which has a similar pattern to other European cities. As can be observed, the POIs of these locations have a greater degree of independence than other cities in the dataset. About 50% of POIs have at least 65% of independence.

Moreover, in the Gowalla dataset it is possible to note that between 30% and 50% of POIs in different cities are completely dependent on their neighborhood (independence = 0) and between 10% and 25% of POIs are completely independent (independence = 1).

A similar analysis can be done in the Weeplaces dataset in Figure 7b. It is possible to note that there are two groups of cities. The first group is formed by U.S. cities and the second group is formed by European cities, however, these groups are not evident as in the Gowalla dataset. Furthermore, U.S. cities have more dependence among POIs and 50% of them have at least 25% of independence, while for European cities POIs are slightly more independent and 50% of them have at least 35% of independence. Also, between 30 and 50% of POIs are completely dependent on their neighborhood and between 10% and 20% are completely independent.

Different from the others, the Twitter dataset in Figure 7c contains U.S. cities only. These cities have a greater degree of independence from other POIs, around 50% of them have at least 70% of independence. The city of San Francisco contains the most dependent POIs among the analyzed cities. Possibly because it is the most touristic city of the dataset, therefore people tend to have longer trajectories among POIs. In contrast, the city of San Antonio has the highest degree of independence among analyzed cities, around 50% of POIs are visited regardless of their neighborhood. Moreover, it is important to note that the Twitter dataset has a larger number of independent locations than other datasets for the same city. This divergent behavior can be explained by the greater number of unique POIs and short trajectories with recurring visits to the same POIs when compared to other datasets. However, as can be seen in Figure 8, our metrics captured almost the same independent regions in the same cities from different datasets.

Therefore, another way to analyze the *independence power* is by means of identification of city areas with greater independence among POIs. To tackle this problem, we divided the cities into geohash [1] cells with size 1.22km x 0.61km. Then, for each cell, we sum the POIs independence. Finally, for each city, we normalized the grid distribution to values between 0 and 1 using the max value and plotted over the city map. Figure 8 shows in green color the most independent areas. It is possible to note that these areas follow a pattern among different cities, i.e., these points are usually concentrated in central areas of cities or areas with a high number of visitations such as



(a) CDF Gowalla

(b) CDF Weeplaces

(c) CDF Twitter

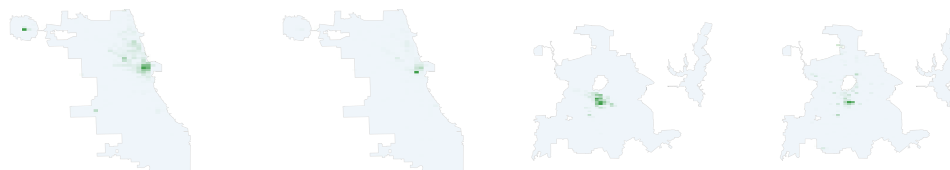
Fig. 7. Most independent locations in the cities

airports or shopping malls. More interesting, our metrics captured almost the same independent regions in the same cities from different datasets.

6 PRACTICAL CONSIDERATIONS AND PERSPECTIVES

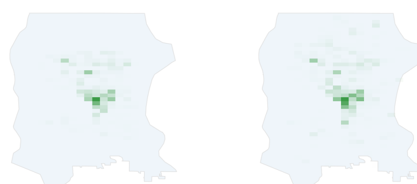
As a practical application of our metrics, we propose to analyze them in a case study of epidemic dissemination. In this case study, our goal is to evaluate how much the neighborhood visitation is affected by the top POIs of each power metric. For this task, first, we exclude from the dataset D POIs with less than 10 visits and, then we grouped the dataset $D = \{d_1, d_2, ..., d_n\}$ with a monthly frequency. Then, for each monthly set d_i we created the *Multiflow Graph Model* and computed the power metrics $\mathcal{M}$. After this step, for each power metric $m_a \in \mathcal{M}$, we listed the top 10 POIs with highest power ($rank_{m_a}$). Thus, for each pair of metrics m_a, m_b, we removed the POI intersection between $rank_{m_a}$ and $rank_{m_b}$, becoming $rank^*_{m_a}$ and $rank^*_{m_b}$, and compared them in the epidemic dissemination scenario. In this scenario, we initially infected the POIs in $rank^*_{m_a}$ and $rank^*_{m_b}$, so that all people who went through one of the contaminated POIs in the period comprising d_i became infected. Finally, at the end of the period comprising d_i, we verified the average percentage of infected people who visited the POIs in m_a and m_b neighborhoods, given by the *Multiflow Graph Model*.

To evaluate the described epidemic scenario, we used the three largest sets of cities from the Gowalla dataset. Moreover, in addition to the previously used metrics, we created a new "joint" metric that adds the values of the *support power* and *attract power* metrics, becoming a single new metric. Figure 9 shows the result of the case study, it is possible to observe a large variance in all

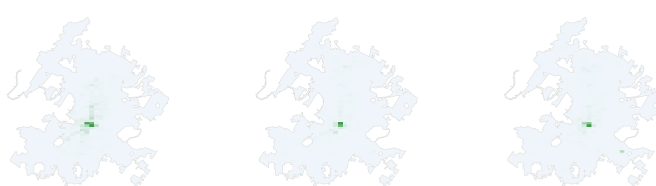


55

(a) Chicago Weeplace (b) Chicago Twitter (c) Dallas Gowalla (d) Dallas Twitter

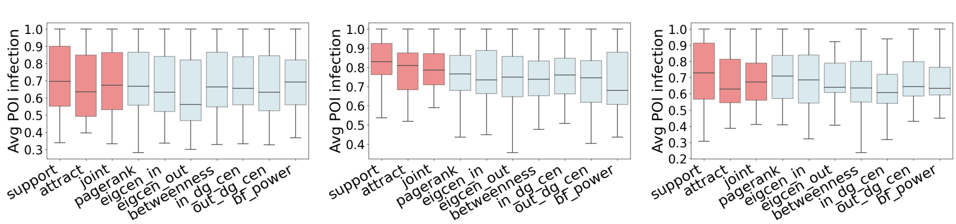


(e) Seattle Gowalla (f) Seattle Weeplace



(g) Austin Gowalla (h) Austin Twitter (i) Austin Weeplace

Fig. 8. Most independent locations in the cities



(a) Austin Gowalla (b) Dallas Gowalla (c) Oslo Gowalla

Fig. 9. Epidemic case study

metrics, however, our metrics of power have a slight advantage, with the median equal to or higher than the other metrics. Highlight to *support power* which is superior to the others, as expected, since we evaluate the dispersion of people around the neighborhood, the main characteristic of this metric. However, in this evaluated context, joining the *attract power* and *support power* metrics in a single metric did not result in a better outcome. The "joint" metric had better results than *attract power* and worse than *support power* alone.

In addition to the previous scenario, we also used the described epidemic scenario to evaluate the POI clusters (high, medium, and low) from Figure 6a in the Dartmouth dataset. The results in Figure 10 show that all clusters had a smaller variance, when compared to the previous scenario. In addition, as expected, the cluster high had a greater impact on its neighborhood with a median of infected people of around 72%. Soon afterward comes the cluster medium with a median of around 65% of infected people, and finally the cluster low with a median of around 23% of infected people. Unlike the clusters high and medium, the cluster low has small *support power* and *attract power* (< 2) and large *independence power* (> 0.6), which explains this huge difference for the other clusters.

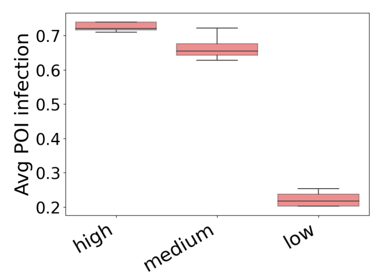


Fig. 10. Cluster epidemic study case

The scenario analyzed above exemplifies an application of our metrics power in a practical scenario of epidemic dissemination. However, we acknowledge that these practical applications need to be deepened to understand the nuances of our metrics, in addition to taking into account other aspects such as dimensional and temporal.

7 CONCLUSION

In this work, we propose a methodology to quantify the power of POIs in three dimensions: *attract power*, *support power*, and *independence power*. We modeled this problem using the *Multiflow Graph Model* where each POI is a node and the transitions of users among POIs are weighted direct edges. The *attract power* and *support power* measure how many visits a POI gather from and disseminate over its neighborhood, respectively. Moreover, the *independence power* is calculated by measuring the number of visits a POI receives that are not influenced by any other POI.

We tested and evaluated our methodology in two groups of public available datasets. The first group, named *test datasets*, describes mobility in three University Campus datasets. The second group, named *validation datasets*, contains users' check-ins/posts in three social networks in different cities around the world.

The results show that, in general, campus buildings have moderate to low *support power* and *attract power*, which is explained by the tendency of people to move over only a few buildings and to spend most of their time in the location. Nevertheless, we identified a slight dependence among buildings, even those buildings with high *independence power* receive user visits from other buildings on campus. Thereafter, we show through counterfactual theories of causation of the form *"If A varies, B also varies"*, that our metrics capture places that impact the number of visits in their neighborhood.

When we split the data by periods of the day, we find that, in some cities, the powerful locations change. Moreover, when investigating the POI independence in the validation dataset, we found that cities on the same continent have similar independence patterns and, in general, central areas and places with a high number of visitations are the regions with the highest degree of independence among POIs. These results show diversity and similarities in patterns between cities that deserve to be investigated and deepened.

The novelty of our approach in methodological terms stems from the use of a graph-based approach, combined with the theory of power relations in exchange networks to tackle human mobility challenges. Moreover, our approach differs from traditional metrics of centrality in some circumstances such as exchange networks where the relationship in the network involves the transfer of valued items (i.e., information, time, money, energy). In this scenario, traditional metrics have limited utility in predicting powerful places.

One limitation of our work is that we only consider the homogeneous impact of POIs in their vicinity, which does not take into consideration the heterogeneity of POIs categories and spatio-temporal bias effects. In future work, we will address these limitations and we will expand our analysis of causality to investigate the POI impact on the neighborhood over time.

REFERENCES

[1] [n.d.]. Geohash. http://geohash.org. Accessed: 2020-02-12.

[2] Licia Amichi, Aline Carneiro Viana, Mark Crovella, and Antonio A F Loureiro. 2019. Mobility profiling: Identifying scouters in the crowd. In *CoNEXT '19: Proceedings of the 15th International Conference on emerging Networking EXperiments and Technologies*. Orlando, United States, 9–11. https://doi.org/10.1145/3360468.3366771

[3] Jordan Bakerman, Karl Pazdernik, Alyson Wilson, Geoffrey Fairchild, and Rian Bahran. 2018. Twitter Geolocation: A Hybrid Approach. *ACM Trans. Knowl. Discov. Data* 12, 3, Article 34 (March 2018), 17 pages. https://doi.org/10.1145/3178112

[4] Mohsen Bayati, Christian Borgs, Jennifer Chayes, Yash Kanoria, and Andrea Montanari. 2015. Bargaining dynamics in exchange networks. *Journal of Economic Theory* 156 (2015), 417–454.

[5] Loris Belcastro, Fabrizio Marozzo, Domenico Talia, and Paolo Trunfio. 2018. G-RoI: Automatic Region-of-Interest Detection Driven by Geotagged Social Media Data. 12, 3, Article 27 (Jan. 2018), 22 pages. https://doi.org/10.1145/3154411

[6] Johannes Bendler, Sebastian Wagner, Tobias Brandt, and Dirk Neumann. 2014. Taming Uncertainty in Big Data. *Business & Information Systems Engineering* 6, 5 (2014), 279–288. https://doi.org/10.1007/s12599-014-0342-4

[7] Geoff Boeing. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65 (2017), 126–139.

[8] Phillip Bonacich. 1987. Power and centrality: A family of measures. *American journal of sociology* 92, 5 (1987), 1170–1182.

[9] Enrico Bozzo and Massimo Franceschet. 2016. A theory on power in networks. *Commun. ACM* 59, 11 (10 2016), 75–83. https://doi.org/10.1145/2934665 arXiv:arXiv:1510.08332v2

[10] Jordan Cambe, Krittika D Silva, E N S De Lyon, Cecilia Mascolo, and Systemes Complexes. 2019. Modelling Cooperation and Competition in Urban Retail Ecosystems with Complex Network Metrics. (2019).

[11] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1082–1090.

[12] Cisco. 2018. Cisco Visual Networking Index: Forecast and Trends, 2017-2022, White Paper.

[13] Cisco. 2018. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper.

[14] Karen S Cook and Toshio Yamagishi. 1992. Power in exchange networks: A power-dependence formulation. *Social networks* 14, 3-4 (1992), 245–265.

[15] Marco De Nadai, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri. 2016. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 413–423. https://doi.org/10.1145/2872427.2883084

[16] Krittika D'Silva, Kasthuri Jayarajah, Anastasios Noulas, Cecilia Mascolo, and Archan Misra. 2018. The Role of Urban Mobility in Retail Business Survival. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–22. https://doi.org/10.1145/3264910

[17] Krittika D'Silva, Anastasios Noulas, Mirco Musolesi, Cecilia Mascolo, and Max Sklar. 2017. If I Build It, Will They Come?: Predicting New Venue Visitation Patterns Through Mobility Data. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Redondo Beach, CA, USA) *(SIGSPATIAL '17)*.

ACM, New York, NY, USA, Article 54, 4 pages. https://doi.org/10.1145/3139958.3140035

[18] Krittika D'Silva, Anastasios Noulas, Mirco Musolesi, Cecilia Mascolo, and Max Sklar. 2018. Predicting the temporal activity patterns of new venues. *EPJ Data Science* 7, 1 (2018), 13. https://doi.org/10.1140/epjds/s13688-018-0142-z

[19] Richard M. Emerson. 1962. Power-Dependence Relations. *American Sociological Review* 27, 1 (1962), 31–41. http://www.jstor.org/stable/2089716

[20] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. DeepMove: Predicting Human Mobility with Attentional Recurrent Networks. *Www* 2 (2018), 10. https://doi.org/10.1145/3178876.3186058

[21] Linton C Freeman. 1978. Centrality in social networks conceptual clarification. *Social networks* 1, 3 (1978), 215–239.

[22] Tristan Henderson, David Kotz, and Ilya Abyzov. 2004. The Changing Usage of a Mature Campus-wide Wireless Network. In *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking* (Philadelphia, PA, USA) *(MobiCom '04)*. ACM, New York, NY, USA, 187–201. https://doi.org/10.1145/1023720.1023739

[23] Ramon Hermoso, Jürgen Dunkel, and Florian Rückauf. 2018. Should I Stay or Should I Go?: Exploiting Visitor Movements to Derive Individualized Recommendations in Museums. In *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (Montreal, QC, Canada) *(MSWIM '18)*. ACM, New York, NY, USA, 343–346. https://doi.org/10.1145/3242102.3242140

[24] Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, Margaret Martonosi, James Rowland, and Alexander Varshavsky. 2011. Identifying Important Places in People's Lives from Cellular Network Data. In *Proceedings of the 9th International Conference on Pervasive Computing* (San Francisco, USA) *(Pervasive'11)*. Springer-Verlag, Berlin, Heidelberg, 133–151. http://dl.acm.org/citation.cfm?id=2021975.2021988

[25] Wei jen Hsu and Ahmed Helmy. 2008. CRAWDAD dataset usc/mobilib (v. 2008-07-24). Downloaded from https://crawdad.org/usc/mobilib/20080724. https://doi.org/10.15783/C79W25

[26] David Kotz and Kobby Essien. 2005. Analysis of a Campus-wide Wireless Network. *Wireless Networks* 11, 1–2 (1 2005), 115–133. https://doi.org/10.1007/s11276-004-4750-0

[27] David Kotz, Tristan Henderson, Ilya Abyzov, and Jihwang Yeo. 2004. CRAWDAD dataset dartmouth/campus (v. 2004-11-09). Downloaded from https://crawdad.org/dartmouth/campus/20041109. https://doi.org/10.15783/C71593

[28] Robert E Kraut, Ronald E Rice, Colleen Cool, and Robert S Fish. 1998. Varieties of social influence: The role of utility and norms in the success of a new communication medium. *Organization Science* 9, 4 (1998), 437–453.

[29] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 739–748.

[30] Peter Menzies. 2017. Counterfactual Theories of Causation. In *The Stanford Encyclopedia of Philosophy* (winter 2017 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[31] John F. Nash. 1950. The Bargaining Problem. *Econometrica* 18, 2 (1950), 155–162. http://www.jstor.org/stable/1907266

[32] Lucas Santos de Oliveira, Pedro O. S. Vaz de Melo, and Aline Carneiro Viana. 2019. Measuring Power Relations Among Locations From Mobility Data. In *Proceedings of the 17th ACM International Symposium on Mobility Management and Wireless Access* (Miami Beach, FL, USA) *(MobiWac '19)*. Association for Computing Machinery, New York, NY, USA, 41–48. https://doi.org/10.1145/3345770.3356744

[33] Ljubica Pajevic, Gunnar Karlsson, and Viktoria Fodor. 2019. CRAWDAD dataset kth/campus (v. 2019-07-01). Downloaded from https://crawdad.org/kth/campus/20190701. https://doi.org/10.15783/c7-5r6x-4b46

[34] Marco Pavan, Stefano Mizzaro, Ivan Scagnetto, and Andrea Beggiato. 2015. Finding Important Locations: A Feature-Based Approach. *Proceedings - IEEE International Conference on Mobile Data Management* 1 (2015), 110–115. https://doi.org/10.1109/MDM.2015.11

[35] Sancheng Peng, Aimin Yang, Lihong Cao, Shui Yu, and Dongqing Xie. 2017. Social influence modeling using information theory in mobile social networks. *Information Sciences* 379 (2017), 146–159. https://doi.org/10.1016/j.ins.2016.08.023

[36] Sharon C Rochford. 1984. Symmetrically pairwise-bargained allocations in an assignment market. *Journal of Economic Theory* 34, 2 (1984), 262–281.

[37] Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham. 2012. Finding Your Friends and Following Them to Where You Are. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (Seattle, Washington, USA) *(WSDM '12)*. ACM, New York, NY, USA, 723–732. https://doi.org/10.1145/2124295.2124380

[38] Richard P Sargent, Robert M Shepard, and Stanton A Glantz. 2004. Reduced incidence of admissions for myocardial infarction associated with public smoking ban: before and after study. *Bmj* 328, 7446 (2004), 977–980.

[39] Ian Shrier and Robert W. Platt. 2008. Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology* 8 (2008), 1–15. https://doi.org/10.1186/1471-2288-8-70

[40] Thiago H. Silva, Aline Carneiro Viana, Fabrício Benevenuto, Leandro Villas, Juliana Salles, Antonio Loureiro, and Daniele Quercia. 2019. Urban Computing Leveraging Location-Based Social Network Data: A Survey. *ACM Comput. Surv.* 52, 1, Article 17 (Feb. 2019), 39 pages. https://doi.org/10.1145/3301284

[41] Lucas M. Silveira, Jussara M. de Almeida, Humberto T. Marques-Neto, Carlos Sarraute, and Artur Ziviani. 2016. MobHet: Predicting human mobility using heterogeneous data sources. *Computer Communications* 95 (2016), 54–68. https://doi.org/10.1016/j.comcom.2016.04.013

[42] Dan Tasse, Zichen Liu, Alex Sciuto, and Jason I Hong. 2017. State of the Geotags: Motivations and Recent Changes. (2017), 250–259.

[43] Douglas Do Couto Teixeira, Aline Carneiro Viana, Mário S. Alvim, and Jussara M. Almeida. 2019. Deciphering Predictability Limits in Human Mobility. In *ACM SIGSPATIAL 2019 - 27th International Conference on Advances in Geographic Information Systems*. Chicago, United States. https://hal.inria.fr/hal-02286128

[44] Waldo R Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46, sup1 (1970), 234–240.

[45] Lei Wang, Ram Gopal, Ramesh Shankar, and Joseph Pancras. 2015. On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems* 76 (2015), 3–13. https://doi.org/10.1016/j.dss.2015.04.010

[46] Liang Wang, Zhiwen Yu, Bin Guo, Tao Ku, and Fei Yi. 2017. Moving Destination Prediction Using Sparse Dataset: A Mobility Gradient Descent Approach. *ACM Trans. Knowl. Discov. Data* 11, 3, Article 37 (April 2017), 33 pages. https://doi.org/10.1145/3051128

[47] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. 2011. Exploiting Geographical Influence for Collaborative Point-of-interest Recommendation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Beijing, China) *(SIGIR '11)*. ACM, New York, NY, USA, 325–334. https://doi.org/10.1145/2009916.2009962

[48] Jia-Dong Zhang, Chi-Yin Chow, and Yanhua Li. 2014. LORE: Exploiting Sequential Influence for Location Recommendations. In *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Dallas, Texas) *(SIGSPATIAL '14)*. ACM, New York, NY, USA, 103–112. https://doi.org/10.1145/2666310.2666400

[49] Xiaoquan Michael Zhang and Feng Zhu. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* 101, 4 (2011), 1601–15.